APPENDIX   F

PREMEETING COMMENTS

Workshop on Selecting Input Distributions for Probabilistic Assessments

Premeeting Comments

New York, New York
April 21-22, 1998

Compiled by:
Eastern Research Group, Inc.
110 Hartwell Avenue
Lexington, MA 02173

❦ERG

Table of Contents

Reviewer Comments

COMMENTS ON THE ISSUE PAPERS / DISCUSSION ISSUES FOR THE EPA WORKSHOP ON SELECTING INPUT DISTRIBUTIONS FOR PROBABILISTIC ASSESSMENT

Probabilistic analysis techniques are, as stated in EPA's May 1997 "Guiding Principles for Monte Carlo Analysis", viable tools in the risk assessment process provided they are supported by adequate data and credible assumptions. In this context, the risk assessor (or risk assessment reviewer) needs to be sensitive to the real-life implications on the receptors of site-specific decisions based on the analysis of variability and uncertainty. The focus should be on the site, in a holistic manner, and all components of the risk assessment should be recognized as tools and techniques used to arrive at appropriate site-specific decisions.

Preliminary (generalized) comments from a risk assessment perspective on the issue papers are provided below, as requested.

Evaluating Representativeness of Exposure Factors Data (Issue Paper #1)

1) The Issue Paper (Framework/ Checklists):

Overall, the issue paper provides a structured framework for a systematic approach for characterizing and evaluating the representativeness of exposure data. However, one of the clarifications that could be provided (in the narrative, checklists and figure) relates to the explicit delineation of the objectives of the exercise of evaluating data representativeness. The purpose of the original study should also be evaluated in the context of the population of concern. In other words, factoring the Data Quality Objectives (DQOs) and the Data Quality Assessment (DQA) premises into the process could help define decision performance requirements. It could also help to evaluate sampling design performance over a wide range of possible outcomes, and address the necessity for multi-staged assessment of representativeness. As stated in the DQA

F-3

Guidance (1997), data quality (including representativeness) is meaningful only when it relates to the intended use of the data.

On the query related to the tiered approach to ("forward") risk assessment, site-specific screening risk assessments typically tend to be deterministic and have been conducted using conservative default assumptions; the screening level tables provided by certain U.S. EPA regions have to this point also been deterministic. Therefore the utility of the checklists at this type of screening level might be extremely limited. As one progresses through increasing levels of analytical sophistication, the screening numbers generated from probabilistic assessment may require a subset of the checklists to be developed; the specificity of the checklists should be a function of the critical exposure parameters identified through a sensitivity analysis. Such analyses might also help refine the protocol (criteria and hierarchy) for assessing data set representativeness in the event of overlap of the individual, population and temporal characteristics (example, inhalation activity in elementary school students in the Columbus area exposed to contaminants at a school ballfield).

2) Sensitivity:
The utility of a sensitivity analysis cannot be overemphasized. Currently, there appears to be a tendency to use readily available software to generate these analyses; guidance on this in the context of project/ site-specific risk assessments should be provided. Providing examples as done in the Region VIII guidance on Monte Carlo simulations facilitates the process.

On the issue of representativeness in making inferences from a sample to a population and the ambiguity of the term "representative sample", process-driven selection might be appropriate for homogenous populations, but for the risk assessor, sampling that captures the characteristics of the population might be more relevant in the context of

the use of the data. This issue appears to have been captured in the discussion on attempting to improve representativeness.

Empirical Distribution Functions (EDFs) versus Parametric Distributions (PDFs) (Issue Paper #2)

1) Selection of the Empirical Distribution Functions (EDF) or Parametric Distribution Function (PDF):

The focus of the issue paper is the Empirical Distribution Function (EDF), and a number of assumptions have been made to focus the discussion on EDFs. However, for a clearer understanding of the issues and to facilitate the appropriate choice of analytical approaches, a discussion of the PDF, specifically the advantages/ disadvantages and constraining situations would be beneficial. The rationale for this is that the decision on whether to apply the EDF or the PDF should not be a question of choice or even mutual exclusivity, but a sequential process that is flexible enough to evaluate the merits and demerits of both approaches in the context of the data.

In general, from a site/ project perspective, there may be definite advantages to PDFs when the data are limited, provided the fit of the theoretical distribution to the data is good, and there is a theoretical or mechanistic basis supporting the chosen parametric distribution. The advantages to the PDF approach are more fully discussed in several references (Law and Kelton 1991). These advantages need to be evaluated in a project-specific context; they could include the compact representation of observations/ data, and the capacity to extrapolate beyond the range of observed data, as well as the "smoothing out" of data.   (In contrast, the disadvantages imposed by the possible distortion of information in the fitting process should not be overtooked. Further, the (traditional use of) EDFs that limit extrapolation beyond the extreme data points, perhaps underestimating the probability of an extreme event, may need to be considered. This is could be a handicap in certain situations, where the risk

assessment demands an interest in outlier values. In such situations, a fuller discussion of alternate approaches such as a mixed-distribution (Brately *et al.,* 1987) may be warranted.) Finally, the PDFs, given their already established theoretical basis, may lend themselves to more defensible and credible decision-making, particularly at contentious sites.

This predisposition to PDFs certainly does not preclude the evaluation of the EDF in the process. The advantage accruing from having the data "speak" to the risk assessor/ reviewer should not be minimized. Depending on the project/ site involved, the benefits of the complete representation of data, the direct information provided on the shape of the underlying distribution, and even on peculiarities such as outlier values should be discussed, as well as relevant drawbacks (sensitivity to random occurrences, potential underestimation of the probability of extreme events, perhaps cumbersome nature if the data points are individually represented). In this context, some of the comments in the "Issue/ Comments" Table ("issues" presumably derived from D'Agostino and Stephens, 1986) can serve as the basis for additional discussion.

2) Goodness of Fit:

The decision whether the data are adequately represented by a fitted theoretical distribution is an aggregative process, and goodness-of-fit is part of the sequential exercise. Preliminary assessments of the general families of distributions that appear to best match the data (based on prior knowledge and exploratory data analysis) are often conducted initially; the mechanistic process for choice of a distributional family, the discrete/continuous and bounded/ unbounded nature of the variable are evaluated. Summary statistics, including measures of shape are evaluated and the parameters of the (candidate) family are estimated. The goodness-of-fit statistics should factor into the whole process, as should graphical comparisons of the fitted and empirical distributions. Goodness-of-fit tests can be an excellent confirmatory tool for verifying

the chosen distribution, when used in conjunction with statistical measures and probability plots.

However, caution should be exercised in situations where these tests could conceivably lead an analyst to support a distribution that a visual inspection of the data does not support. Also, it should be emphasized that (for example for certain physiological parameters), even if the distribution fits, maintaining the integrity of the (biological) data should override goodness-of-fit considerations. Ultimately, the persuasive power of graphical methods for assessing fit should not be underestimated.

On the question how the level of significance of the goodness-of-fit statistic should be chosen, this is often a function of the data quality assessment (DQA) for that particular site or situation; an idea of the consequences in terms of real-life examples can be gathered from EPA's Guidance for Data Quality Assessment (1997). On the whole, I tend to agree with the respondent (#4) who states that the desired level of significance should be determined prior to analyzing the data. Again, as the respondent states, if minor differences in the p-value impinge substantially on the analysis, the "conclusions are probably too evanescent to have much usefulness".

Summary statistics are useful, particularly in the initial characterization of the data (as previously mentioned). Given the constraints imposed by the project/ site logistics, all too often these are the only data available, and they have been used as the basis for analytical distribution fits (Ohio EPA, 1996). Caution should be exercised in implying a level of accuracy based on limited knowledge. Sensitivity analyses might help clarify the limitations that need to be placed in such situations particularly when dealing with an exposure parameter of considerable impact; further, the utility of such an exercise for a parameter with minor impact (as revealed by the sensitivity analysis) could be questionable.

On the question of the value of testing the fit of the more generalized distributions (presumably in lieu of the EDF), this could be an useful exercise, but the project logistics may factor into this, as also the DQA premises. Project resources available and the defensibility of the decision-making process need to be factored into the situation. The issue of fitting an artificial distribution to a data set, and ultimately arriving at a distribution removed from reality also needs to be evaluated in the project-specific context.

3) Uncertainty:

The discussion in "Development of Statistical Distributions for Exposure Factors" (Research Triangle Institute) paper is interesting in terms of the approaches suggested for evaluating parameter uncertainty; Hattis and Bummaster's comment cited in the paper that only a trivial proportion of the overall uncertainty may be revealed is important. Certain methods (example, bootstrapping) appear to have intriguing potential for accounting for "hot spots".

Finally, the risk assessor/ reviewer needs to be aware that the analysis of variability and uncertainty is a simulation, based on hypothetical receptors. However, as stated initially, this sometimes academic exercise can have multi-million dollar implications, and intimately affect real-life human and ecological receptors; the risk assessor/ reviewer should always be cognizant of this consequence.

References:

Brately, P., B.L. Fox, L.E. Schrage (1987) "A Guide to Simulation". Springer-Verlag, New York.

D'Agostino, R.B. and M.B. Stevens (1986) "Goodness of Fit Techniques". Marcel Deker.

Law, A.M. and Kelton, W.D. (1991) "Simulation Modeling and Analysis" (Chapter 6, 325-419). McGraw-Hill, New York.

Ohio EPA (1996) "Support Document for the Development of Generic Numerical Standards and Risk Assessment Procedures". The Voluntary Action Program, Division of Emergency and Remedial Response, Ohio EPA.

U.S. EPA (1994) "Guidance for the Data Quality Objectives Process" (EPA/QA/G4). EPA/600/R-96-055

U.S. EPA (1997) "Guidance for Data Quality Assessments - Practical Methods for Data Analysis" (EPA QA/G-9, QA-97 Version) EPA/600/R-96/084 (January 1998)

Robert J. Blaisdell, Ph.D.

Comments on Issue Paper on Evaluating Representativeness of Exposure Factors Data

The Issue Paper on Evaluating Representativeness of Exposure Factors Data is a well written, clear discussion of the theoretical issues of representativeness. I was particularly interested in the discussion of time unit differences. The Office of Environmental Health Hazard Assessment (OEHHA) is grappling with this issue with several of the distributions which we want to use for determining chronic exposure.

The issue of representativeness of a sample is often complicated by lack of knowledge about the demographics of the population under consideration. An accurate determination of the population under consideration may not be part of the risk assessment requirements of regulatory programs. If the population of concern has not been characterized, the determination of the representativeness of the data being used in the assessment is not possible.

The issue of representativeness of the sample to the population is an important question. For example, populations which are exposed to Super Fund toxicants or airborne pollution from stationary sources may be from lower socioeconomic groups. Unfortunately, most of the information which is available **on** mobility is from the general population. It may be that low income home owners have a much longer residency time than people of median or higher income. It may also be that low income non-home owners in certain age groups have a higher mobility than the general population. We therefore suspected that the available distributions were not representative. In addition, the U.S. Census data, the basis for the available residency distributions are not longitudinal. Another problem with the residency data when evaluating stationary sources is the issue of where the person moves to. A person moving may not necessarily move out of the isopleth of the facility. The likelihood of moving out of the isopleth of a stationary facility also may be related to socioeconomic status.

In order to address this problem, OEHHA proposed not using a distribution for residence time in our Public Review Draft Exposure Assessment and Stochastic Analysis Technical Support Document (1996). Instead we proposed doing a separate stochastic analysis scenario for 9, 30 and 70 years. We did not think that the 9, 30 or 70 years time points evaluated were necessarily representative of actual residence times, but that these were useful, reasonably spaced intervals for residents to compare with their own known residency time.

Using three scenarios complicates the analysis, but we felt that the approach had some advantages over using a distribution. The California *Hot Spots* program is a public right to know act which assesses risks of airborne pollutants from stationary sources. Public notification is required above a certain level of risk. An individual resident who has received notice is aware of the amount of the time that he or she has lived, or in many cases plans to live, in vicinity of the facility. Therefore the individual could more accurately assess his or her individual cancer risk. The relationship between the residency time assumption and the resulting risk are clear, not buried in the overall range of the uncertainty or variability of the risk estimate.

This approach might possibly be used in other cases where representative data in not available or where the representativeness is questionable. For example if the drinking water pathway is of concern and representative information is not available for the population of a Mojave Desert town, the range or point estimate of cancer risk from drinking 1, 2, 4 and 8 liters of contaminated tap water per day could be presented.

In some cases, each situation that a regulatory risk assessment program will be evaluating will be almost unique, and therefore anything other than site-specific data will not be representative. OEHHA characterized a fish consumption distribution for anglers consuming non-commercial fish using, the Santa Monica Bay Seafood Consumption Study Final Report (6/94) raw data. We compared the Santa Monica Bay distribution to

the fish consumption distribution for the Great Lakes (Murray and Butmaster, 1994). We found that the differences in the two distributions could be attributed to methodological differences in the two studies. Thus the assumption that a salt water fish consumption distribution was comparable to a fish consumption distribution for large fresh water body was not implausible. However, the data gathered from large bodies of water are probably not representative of small lakes and ponds with limited productivity and where other fishing options may exist. For such bodies of water a site-specific angler survey is probably the only way of obtaining representative data. For cost reasons, this option is not likely to be pursued except in a risk assessment with very high financial stakes. We chose to recommend using the Santa Monica Bay fish consumption. It could be multiplied by a fraction to be determined by expert judgment to adjust for site-specific conditions such as productivity etc. The Santa Monica Bay fish distribution may not be representative in other ways in a given situation but may still be the most practical option. It is clearly not temporally representative for chronic cancer risk assessment.

Cost is often a factor that limits representativeness.

On page 8, paragraph 3 of the Issues paper there is a discussion of determining the relationship between two populations and making adjustments in distributions based on speculative estimates of the differences in means and the coefficients of variation. Perhaps in many instances, another option would be to state that the information from a surrogate population is being used and that the actual population is known to be different, or may be different by an unknown amount. There are many questions in risk assessment for which expert opinion is no better than uninformed opinion in attempting to quantify the unknown. An example of this is the shape of the dose-response curve for cancer for most chemicals at low concentrations.. A frank admission of ignorance may be more credible than an attempted quantification of ignorance in many cases,

Robert J. Blaisdell, Ph.D.

Comments on Temporal Issues

The methods discussed for estimating intraindividual variability from data collected over varying short periods of time relative to the longer time period of interest are interesting and would appear to be useful for the NFCS data. OEHHA is giving some consideration to using the techniques described by Nusser *et al.* 1996 to adjust the distributions for food consumption that we have developed for food consumption using the Continuing Survey for Food Intake for Individuals 1989-91 raw data. I would be curious to know if these methods have been validated on any actual longitudinal data. The assumption of the lognormal model needed by the method of Wallace et al. (1994) may in some cases be limiting. We have discovered when we evaluated broad categories of produce consumption using the CSFII 89-91 data that some of the distributions for certain age groups were closer to a normal model than a lognormal model.

The Representativeness Issue paper discusses the importance of using current data. The continued use of the 1977-78 NFCS study is cited as an example. The raw data from the 1989-91 CSFII has been available for some time as an alternative to the 1977-78 NFCS survey. Raw data from the 1992-93 CSFII survey is now available. OEHHA has used that data to develop produce, meat and dairy products consumption distributions for the California population. It is admittedly not a trivial exercise to extract the relevant data from the huge raw CSFII data sets but this alternative has existed for several years. The 1989-91 CSFII data is clearly different in some cases from the 1977-78 NFCS. Beef consumption appears to have declined. As a matter of policy, there should be a stated preference for using the available data over attempting to use expert judgment to guess at the appropriate means, coefficients of variation and parametric model. In some of the Monte Carlo risk assessment literature, the preference appears to be for expert judgment rather than data.

The use of related data may in some cases be useful in giving some insight into the representativeness of data collected over the short term for chronic scenarios. OEHHA has used the data on total energy expenditure as measured by the doubly labeled water method to look at the representativeness of our breathing rate distribution, based in part on a one day 24 hour activity pattern survey. The information on total energy expenditure gave an indication that intraindividual variability was a huge fraction of the total variability (intraindividual plus interindividual variability).

The intraindividual variability for a broad category of produce such as leafy vegetables may not be very great relative to the interindividual variability. The intraindividual variability for a single item less frequently consumed item such as strawberries is probably much greater than for broad categories. Thus, short term survey data which looks at broader categories of produce are probably more applicable to chronic risk assessment than single item distributions.

Research Needs

The information which is needed to develop more accurate distributions for many if not most variates needed for chronic stochastic human health risk assessment are simply not available. In particular there is a lack of longitudinal data for breathing rates, soil ingestion, water consumption rates, produce ingestion, non-commercial fish consumption, dairy product consumption and meat ingestion. Some distributions in common use, such as water consumption, are based on out of date studies. More research is needed on bioconcentration and biotransfer factors. Longitudinal data on activity patterns and mobility patterns would also be very useful. There needs to be much more research on dermal absorption factors and factors which influence dermal absorption. More research needs to be done on children and the ways that they differ from adults.

Summary

The overall lack of data, particularly longitudinal data, for risk assessment variates is probably the most important single factor limiting representativeness. If the purpose of the risk assessment is to inform the exposed public, it may be possible and even preferable to use point estimates for multiple scenarios in the absence of some representative data. The statistical methods for adopting short term data for use in chronic risk assessment presented the Issue paper appear to be reasonable approaches in instances where the required data is available. More longitudinal studies would be valuable for validation of these methods as well as improving the temporal representativeness of distributions used in risk assessment. Most of the data used in stochastic risk assessment will probably be nonrepresentative in one or more of the ways discussed in the Issues paper for a long time into the future.

References

Murray DM., and Burmaster DE. (1994). Estimated distribution for average daily consumption of total and self-caught fish for adults in Michigan angler households. Risk Analysis 14, 513-519.

Nusser, S.M., Carriquiry, A. L., Dodd, D.W., and Fuller, W. A. A semiparametric transformation approach to estimating usual daily intake distributions. J. Am. Statistical Association 91: 1440-1449, 96.

Southern California Coastal Water Research Project and MBC Applied Environmental Sciences (SCCWRP and MBC). (1994). Santa Monica Bay Seafood Consumption Study. Final Report. June.

Robert J. Blaisdell, Ph.D.

USDA (U.S. Department of Agriculture) 1989-91. Nationwide Food Consumption Survey. Continuing Survey of Food Intakes of Individuals (Data Tapes) Hyattsville, Md: Nutrition Monitoring Division, Human Nutrition Information Service.

David Burmaster

13 April 1998

Memorandum

To: Participants, US EPA's Workshop on Selecting Input Distributions
                for Probabilistic Analyses

Via: Beth A. O'Connor, ERG

From: David E. Burmaster

Subject: Initial Thoughts and Comments,
                and Additional Topics for Discussion

Thank you for inviting me to participate in this Workshop in New York City.

Here are my initial thoughts and comments, along with suggestions for additional topics for discussion. Since I have just returned from 3 weeks of travel overseas, I will keep these brief.

1.   Models and Data

In 1979, George Box wrote, "All models are wrong, but some are useful."

May I propose a new corollary for discussion? "All data are wrong, but some are useful."

2.   Definitions for Variability and Uncertainty

The Issue Papers lack crisp definitions for <u>variability</u> and <u>uncertainty</u> as well as a discussion about why variability and uncertainty are important considerations in risk assessment and risk management. (See, for example, NCRP, 1996.) In particular, I recommend definitions along these lines for these two key terms:

- <u>Variability</u> represents true heterogeneity in the biochemistry or physiology (e.g., body weight) or behavior (e.g., time spent showering) in a population which cannot be reduced through further measurement or study (although such heterogeneity may be disaggregated into different components associated with different subgroups in the population). For example, different children in a population ingest different amounts of tap water each day. Thus variability is a fundamental <u>property of the exposed population</u> and or the exposure scenario(s) in the assessment. Variability in a population is best analyzed and modeled in terms of a full probability distribution, usually a first-order parametric distribution with constant parameters.

- <u>Uncertainty</u> represents ignorance -- or lack of perfect knowledge -- about a phenomenon for a population as a whole or for an individual in a population which may sometimes be reduced through further measurement or study. For example, although we may not know much about the issue now, we may learn more about certain people's ingestion of whole fish through suitable measurements or questionnaires. In contrast, through measurements today, we cannot now eliminate our uncertainty about the number of children who will play in a new park scheduled for construction in 2001. Thus, uncertainty is a <u>property of the analyst</u> performing the risk assessment. Uncertainty about the variability in a population can be well analyzed and modeled in terms of a full probability

distribution, usually a second-order parametric distribution with nonconstant (distributional) parameters.

Second-order random variables (Burmaster & Wilson, 1996; references therein) provide a powerful method to quantify and propagate V and U separately.

3.   Positive Incentives to Collect New Data and Develop New Methods

I urge the Agency print this Notice inside the front cover and inside the rear cover of each Issue Paper / Handbook / Guidance Manual, etc. related to probabilistic analyses — and on the first Web page housing the electronic version of the Issue Paper / Handbook / Guidance Manual:

<div style="border:1px solid black; padding:10px;">

This Issue Paper / Handbook / Guidance Manual contains guidelines and suggestions for use in probabilistic exposure assessments.

Given the breadth and depth of probabilistic methods and statistics, and given the rapid development of new probabilistic methods, the Agency cannot list all the possible techniques that a risk assessor may use for a particular assessment.

The US EPA emphatically encourages the development and application of new methods in exposure assessments and the collection of new data for exposure assessments, and nothing in this issue Paper / Handbook / Guidance Manual can or should be construed as limiting the development or application of new methods and/or the collection of new data whose power and sophistication may rival, improve, or exceed the guidelines contained in this Issue Paper / Handbook / Guidance Manual .

</div>

4.   Truncating the Tails of LogNormal Distributions

While LogNormal distributions provide excellent fits to the data for many exposure variables, e.g., body weight, skin area, drinking water ingestion rate (total and tap), showering time, and others, it is important to truncate the tails of these distributions. For example, no individual has 1 cm^2 of skin area; no individual has 10^5 cm^2 of skin area, and no individual can shower 25 hr/d.

5.   Mixing Apples and Oranges

It is wholly inconsistent for the Agency to proceed with policies that legitimize the use of probabilistic techniques for exposure factors while preventing the use of probabilistic techniques in dose-response assessment. By doing so, the Agency double counts the effects of variability and uncertainty, all on a $\log_{10}$ scale -- i.e., by several orders of magnitude.

6.   Report by RTI

I disagree strongly with many of the approaches and conclusions found in RTI's Final Report dated 18 March 1998.

References

Box, 1979

Box, G.E.P., 1979, Robustness is the Strategy of Scientific Model Building, in Robustness in Statistics, R.L Launer and G.N. Wilkinson, eds., Academic Press, New York, NY

Burmaster & Wilson, 1996

But-master, D.E. and A.M. Wilson, 1996, An Introduction to Second-Order Random Variables in Human Health Risk Assessment, Human and Ecological Risk Assessment, Volume 2, Number 4, pp 892 - 919

REPRESENTATIVENESS (Issue Paper #1)

1) The Issue Paper

We would use probabilistic methods specifically for the purpose of assessing risks from the uncontrolled release of hazardous substances at a specific location (site). Our overall goal will be to feel confident that the entire risk assessment (and not just a few of its components) is representative of site-specific conditions. Our objective is better risk management decisions. This requires us to keep a few other considerations in mind.

The issue of representativeness in terms of a fit between available exposure factors data and resulting distributions is dealt with in the issue paper. However, a risk assessment cannot be performed with exposure factor distributions alone - some type of exposure model is required. We should therefore also be concerned with the representativeness of the exposure model within which the individual exposure factors are used.

Correlation between exposure factors could significantly affect the representativeness of the resulting risk assessment. It appears possible to have too much or little correlation between factors. In some cases, the correlation is not necessarily with body weight and/or age but with an underlying activity pattern (human behavior) that may not be fully known. This nature and extent of correlation should be a factor in evaluating representativeness.

The issue of data and statistical inferences at the extreme upper bounds (e.g., 99.9th percentile) of a distribution has been raised in the literature, on the Web, and in other U. S. EPA forums. As a matter of policy, we regulate at the 90th percentile, feel that decisions based on extreme upper bound estimates are potentially unreasonable, and thus have truncated the upper bound (not allowed its extension to +^{oo}) of many of

F-22

the exposure factor distributions. How any such truncation of a distribution affects its representativeness should also be discussed.

The suggestion that probabilistic methods could be used in any form of "screening-level" risk assessment is of concern. We view screening has a quick but highly conservative comparison of environmental media concentrations with published toxicity data that occurs early in a remedial investigation (RI) for the sole purposes of narrowing the focus of the baseline risk assessment. Under our current guidance, we are preserving probabilistic methods for use only in a baseline assessment.

2) Sensitivity

When various exposure factors are combined within a given exposure model, it is typically the case that a few of them have a disproportionate influence on the outcome. For example, soil ingestion rate, soil adherence factor, and exposure duration are often primary drivers, as well as major sources of uncertainty. We should broaden the discussion to consider whether all exposure factors are of equal importance, in terms of their influence on the outcome of the risk assessment, so as to better focus our distribution development efforts.

3) Adjustments

Concern has been expressed that any "default" exposure factor distributions proposed by U. S. EPA will, perhaps unintentionally, will evolve into inflexible or "standard" requirements. To counter this, as well as allow for inclusion of regional and local influences, U. S. EPA should propose, in addition to any de facto "default' distributions, an exemplary method(s) for establishing exposure factor distributions. This exemplary method should be as straightforward, transparent, and explainable (primarily to risk managers) as possible. It should also describe quality assurance (QA) and quality control (QC) procedures to allow for the expedient and thorough review of probabilistic risk assessments submitted to regulatory agencies by outside contractors.

Bruce Hope

EMPIRICAL DISTRIBUTION FUNCTIONS (Issue Paper #2)

{I did not have time to fully review paper #2, so only have input on this one item at this time}

2) Goodness of Fit

We should also ask, if the overall risk assessment is sensitive to both the exposure model and only a few of many exposure factors, just how "good" does every other distribution have to be in order to support credible risk management decisions? For example, if a relatively esoteric and hard to conceptualize distribution best tits available data, but a much more common and more easily understood distribution fits almost as well (say within 20%), would there not be some advantage in use of the latter? In addition, if toxicity data remain as point estimates with uncertainty approaching an order-of-magnitude, it would appear that there should be some leeway in how we choose or define certain exposure factors.

William A. Huber

Representativeness (Issue Paper #1)

1)   The Issue Paper

1.1   The checklists

Section 3 of the Issue Paper regards the inferential process as consisting of several stages of inference and measurement: Population of interest -> Population(s) actually studied -> Set of individuals measured (the "sample") -> The measurements. The three stages are denoted "external" inference, "internal" inference, and measurement, respectively.

This appears to be a useful framework. However, the four checklists address the first two stages only. Checklist I concerns the "internal" inference; Checklists II through IV concern the "external" inference. No checklist specifically addresses measurement. This approach is unbalanced. The obvious parallelism among Checklists II through IV emphasizes the lack of balance. We should consider whether a better organization of checklists might be achieved. One possible organization could be:

Checklist A: Assessing measurement representativeness

Checklist B: Assessing internal representativeness

Checklist C: Assessing external representativeness

Checklist D: "Reality checks," or overview.

Checklist B and checklist I would nearly coincide. Checklist C would incorporate the (common) questions of checklists II through IV. Checklists A and D are new. Checklist A would incorporate certain questions sprinkled throughout Checklists I-IV, such as:

● Does the study appear to have and use a valid measurement protocol?

● To what degree was the study design followed during its implementation?